

Bounding the seed length of Miller and Shi's unbounded randomness expansion protocol

Renan Gross and Scott Aaronson

Abstract

Recent randomness expansion protocols have been proposed which are able to generate an unbounded amount of randomness from a finite amount of truly random initial seed. One such protocol, given by Miller and Shi, uses a pair of non-signaling untrusted quantum mechanical devices. These play XOR games with inputs given by the user in order to generate an output. Here we present an analysis of the required seed size, giving explicit upper bounds for the number of initial random bits needed to jump-start the protocol. The bits output from such a protocol are ε -close to uniform even against quantum adversaries. Our analysis yields that for a statistical distance of $\varepsilon = 10^{-1}$ and $\varepsilon = 10^{-6}$ from uniformity, the number of required bits is smaller than 225,000 and 715,000, respectively; in general it grows as $O(\log \frac{1}{\varepsilon})$.

1 Introduction

Building a device that generates a random string using quantum mechanics is easy: All it needs to do is to prepare a qubit in a state in the X basis, and then measure it in the Z basis. However, what if you didn't build the device yourself, but instead it was given to you by your arch-nemesis? How could you certify that the output is indeed random, and not, for example, deterministically fixed, or somehow correlated to the arch-nemesis? To treat these problems, protocols have been developed recently which allow one to certify that the output of an untrusted device was indeed random (R. Colbeck, [2]). The devices in these protocols don't just prepare and measure qubits in different states, but are made of components that play XOR games with each other. As these games require random bits as input, effort was invested in generating more randomness than was invested as input; this is called randomness expansion. Both polynomial (S. Pironi et. al [9]) and exponential (Vazirani and Vidick [11]) expansion has been described, and recently even infinite expansion (Chung, Shi, Wu and Miller [1, 8], Coudron and Yuen [4]). The latter protocols take as input a finite truly random string and output a nearly uniform string of arbitrary length; further, the distance from uniformity depends only on the number of random bits used as seed. All of [1, 4, 8] rely on a "spot checking" technique by Vazirani and Vidick [11] and Coudron, Vidick and Yuen [3] when generating inputs to the XOR games; their contribution and differences are in how they compose the inputs and outputs between devices and the analysis of that composition. The proofs given in those papers are asymptotic and do not give concrete bounds on the number of initial random bits needed in order to provide this infinite expansion with the desired soundness and security. In this paper, we give a rough estimate of the number of bits needed in order to obtain a desired distance from uniformity. We follow the analysis of Miller and Shi [8], who, in conjunction with Chung, Shi, and Wu [1], built a protocol which uses only two devices. This protocol appears to be simpler to analyze than the one suggested by Coudron and Yuen [4] and is likely to have

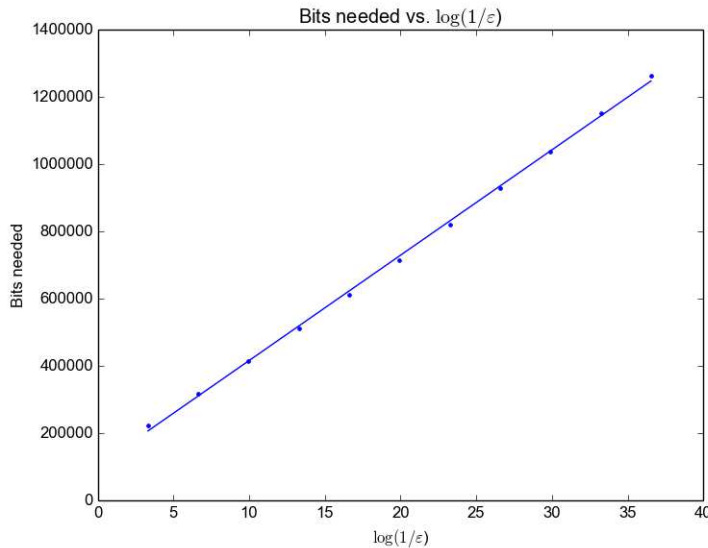


Figure 1: The seed length obtained by using the technique described in Section 4. The linear dependence was obtained by passing a fit through the needed seed length for different ε 's in the range 10^{-1} to 10^{-11} .

smaller constant overhead: it uses less quantum devices and does not use the relatively complicated Reichardt-Unger-Vazirani protocol [10].

The analysis yielded an upper bound for a single iteration of the protocol - which gives exponential expansion - and a technique for numerically approximating the number of bits needed for unbounded iterations. For example, for an error of 10^{-6} , the seed length is bounded from above by 715,000; generally, from numerical calculations, the relation between the seed length approximation and the error can be bounded from above by a linear relation: $\text{SeedLen} \propto 31328 \cdot \log_2 \frac{1}{\varepsilon}$, as can be seen in Figure 1.

Randomness is needed for two purposes: generating XOR games, and extracting randomness from high entropy strings. As we will see later, generating the XOR games is the more demanding of the two, taking the larger portion of the overall random bits needed: the ratio between them is about 2:1.

2 Spot Checking protocol

2.1 The gist of the protocol

The protocol requires two identical non-interacting *devices*. A device consists of n non-interacting *components*; these components are going to play a XOR game. The number n depends on the game being played; thus the CHSH game requires two components, while the GHZ requires components. A single run consists of having a *single* device play a very large number of games. If the device wins enough games, the protocol succeeds and an output is generated according to its answers; by an appropriate variant of the Bell inequality (depending on the game), this output is guaranteed to have some min-entropy. If the device doesn't win enough games, the protocol aborts.

The main point is that the input to the games is not uniformly random; in fact, most of the time, the input is just zeros. Only on a small, randomly selected number of games are the inputs chosen at random. As the device doesn't know where the randomized inputs are going to be, this forces it "play honest" and play a non-deterministic winning strategy on almost all of the games, if it wants to pass the protocol. Thus, a string with high min-entropy can be obtained, while using very little random bits - most of the inputs were predetermined zeros.

The output string can then be fed into a randomness extractor, yielding a nearly-uniform random string. The above procedure is then repeated again and again, each time using the extracted string as a source of randomness for choosing where the non-zero games are, and each time alternating between the two devices (the alternation is an important part of proving the quantum security, but is not needed in our analysis, and we will not go in detail about it here).

2.2 More formally

The protocol is composed of *iterations*. During an iterations a single device is used, and all of its components play the XOR game. The following arguments are fixed:

N : The **output length**. This is a positive integer which denotes how many times we will play the game.

η : The **error tolerance**. This is a real number $\in (0, \frac{1}{2})$ which denotes how large a statistical error we allow our components to make relative to the optimal winning strategy's expectation.

q : The **test probability**. This is a real number $\in (0, 1)$ which denotes the probability that a round will be a randomized "game round" (see ahead).

The single iteration protocol, denoted R , is then as follows:

1. Repeat steps 2-4 N times:
2. A bit $g \in \{0, 1\}$ is chosen according to the distribution $(1 - q, q)$
3. If $g = 1$ ("game round"), then an input string is chosen at random from $\{0, 1\}^n$, according to the specific game chosen. For example, for the GHZ game, the possible strings are 000, 100, 010, 001. If the devices win, record 0. Else, record 1, and mark "Failure".
4. If $g = 0$ ("generation round"), then the input string composed entirely of zeros 00...0 is given to the components. Record the bit generated by the first component.
5. If the total number of failures exceeds $(1 - \mathbf{w}_G + \eta)qN$, where \mathbf{w}_G is the winning probability for the optimal strategy, the protocol aborts. Otherwise it succeeds, and outputs the N -bit sequence of outcomes it recorded.

It can be shown that with the right choice of parameters η, q , and N , the output string can be ε -close to $(1 - \delta)N$ min-entropy for any choice of δ , with ε exponentially small as a function of N . The amount of randomness needed to generate this string goes roughly as $\log N$, so it is possible to generate a string with arbitrarily more min-entropy than what we started with.

A quantum-secure extractor is then applied to the output, yielding a smaller but nearly uniform random string. It is possible to construct extractors that extract a constant fraction of min-entropy, while using an additional seed of size $O(\log^2 \frac{N}{\varepsilon})$, where ε is the distance to uniformity¹.

Thus, running one iteration and applying an extractor yields exponentially many more bits than we started with. By alternating between devices and using the output of one device as the randomness seed for the game generation and extractor of the other, an unbounded amount of random bits can be produced.

2.3 Layout

We start by analyzing the seed length needed for a single iteration of the protocol: given a target error in uniformity, how many random bits do we need in order to get just the exponential expansion for one device? We then look at how the error grows when we play several iterations of the protocol. The XOR game used by the devices has been chosen to be the GHZ game, for several reasons: it features a large gap between the best quantum strategy (100% win rate) and the best classical strategy (75% win rate); its best strategy always wins; and Miller and Shi give a bound to its “trust coefficient”, a constant that appears in their analysis.

3 Single iteration with extraction

A single iteration requires randomness in two places: choosing the inputs, g_i to the XOR game for the device, and seeding the extractor. These two are not quite independent of each other: if we play N games with the device, our output will be a string of length N with min-entropy linear in N . Both the game input randomness, and the extractor seed length are polylogarithmic in N . We will start by analyzing the XOR game, and then proceed to the extractor seed.

3.1 Definitions

We follow the same notation as Miller and Shi. Logarithms written as $\log x$ are in base 2; logarithms written as $\ln x$ are natural.

For a given XOR game G , let \mathbf{f}_G be smallest *failing probability* for a game; that is, the probability that the best quantum strategy will fail to win the given game. For the GHZ game, we have $\mathbf{f}_G = 0$. The *trust coefficient* for a game is a number $\mathbf{v}_G \in (0, 1]$ (described in more detail in the main text, but no more than this is needed). For the GHZ game, it was proven that $\mathbf{v}_G \geq 0.14$; we will denote this bound as $c_v = 0.14$.

The following functions appear in the theorems and corollaries:

$$\Pi(x, y) \triangleq 1 - \left(\frac{1 + 2x}{x} \right) \log \left[(1 - y)^{\frac{1}{1+2x}} + y^{\frac{1}{1+2x}} \right]$$

$$\pi(x) \triangleq 1 + 2x \log x + 2(1 - x) \log x = 1 - 2h(x)$$

It can be shown that $\lim_{x \rightarrow 0} \Pi(x, y) = \pi(y)$. The derivative of the π function is

¹A quantum secure extractor is needed only if we are afraid that an adversary might be entangled with the internal mechanism of the device, and thus gain information about our random string. If this is not the case - if we only wish to verify that the device generates random bits without conditioning on a possible adversary’s information - then a constant fraction extractor can be built with a seed size of only $O(\log \frac{N}{\varepsilon})$ [6].

$$\pi'(y) = 2(\log(1 - y) - \log y)$$

For $y \leq 0.5$ we have:

- $\pi(y)$ is non-increasing and has a minimum of -1 at $y = 0.5$.
- $\pi'(y)$ is negative and non-decreasing, with a zero at $y = 0.5$. It tends to $-\infty$ for $y \rightarrow 0$.

$H_{min}^\varepsilon(S)$ denotes the ε -smooth min-entropy of a quantum state S . The state Γ_{EGIO}^s denotes the state of success of a iteration for a given adversary E , game G , input I and output O .

3.2 Choosing randomness for the XOR game

The corollary numbering in this section is according Miller and Shi's paper [8]. An important result in that paper is **Corollary I.5**, which states:

Corollary (I.5). *Let $\delta > 0$ be a real number. Then, there exists positive reals K, b, q_0 and η such that the following holds. If Protocol R is executed with parameters N, η, q , where $q \leq q_0$, then*

$$H_{min}^\varepsilon(\Gamma_{EGIO}^s | EGI) \geq N \cdot (1 - \delta)$$

where $\varepsilon = K \cdot 2^{-bqN}$.

This is done by proving lower bounds on the rate of entropy. To quote Miller and Shi: “Our approach, broadly stated, is as follows: we show the existence of a function $T(v, h, \eta, q, k)$ which provides a lower bound of the linear rate of entropy of the protocol. [...] In principle, our proofs could be used to compute an explicit formula for the function T , but we have not attempted to do this because the formula might be very complicated.”

In light of these words, they have only shown the behavior of T in the limit of small q and k parameters. In the following section, we will:

- Attempt to find explicit bounds for T for “small enough” parameters.
- Calculate how many random bits are needed, given fixed q and N , in order to obtain desired min-entropy rate and smoothness values.

3.2.1 Bounding the T function

In the original paper, the T function is given by composition of a sequence of other functions. It can be expanded to yield:

$$\begin{aligned} T(v, h, \eta, q, k) \triangleq & -\frac{1}{r_0 q k} \cdot \max_{t \in [0, 1]} \left[\log \left((1 - q) 2^{-r q k \cdot \Pi(r_0 q k, t)} + \right. \right. \\ & \left. \left. + q \left\{ 1 - (1 - 2^{-k}) \left[\left(\frac{h}{2} \right)^{1 + r_0 q k} + v^{1 + r_0 q k} t \right] \right\} \right) \right] - \frac{\frac{h}{2} + \eta}{r_0} \end{aligned}$$

where

$$r_0 \triangleq \min \left\{ \frac{-v}{\pi'(\frac{\eta}{v})}, \frac{1}{qk} \right\}.$$

Effectively, because q and k can be made arbitrarily small, $r_0 = \frac{-v}{\pi'(\frac{\eta}{v})}$. We also define:

$$E(v, h, \eta, q, k) \triangleq \frac{2}{r_0}$$

Which, under our assumption, simplifies to:

$$E(v, h, \eta, q, k) = \frac{-2\pi'(\frac{\eta}{v})}{v}$$

Miller and Shi use these functions to talk about the min-entropy found in the output of a large number of games:

Theorem (I.1). *Suppose Protocol R is executed with parameters N, η, q . Then for any $k \in (0, \infty)$ and $\varepsilon \in (0, \sqrt{2}]$, the following holds:*

$$H_{min}^\varepsilon(\Gamma_{EGO}^s | EG) \geq N \cdot T(\mathbf{v}_G, 2\mathbf{f}_G, \eta, q, k) - \left(\frac{\log(\sqrt{2}/\varepsilon)}{qk} \right) E(\mathbf{v}_G, 2\mathbf{f}_G, \eta, q, k).$$

Also,

$$\begin{aligned} \lim_{(q,k) \rightarrow (0,0)} T(v, h, \eta, q, k) &= \pi\left(\frac{\eta}{v}\right) \\ \lim_{(q,k) \rightarrow (0,0)} E(v, h, \eta, q, k) &= \frac{-2\pi'(\frac{\eta}{v})}{v}. \end{aligned}$$

We will now look at the appropriate theorems from their papers and root out the needed constants from their proofs.

Corollary (I.2). *for every $\eta > 0$ and $\delta > 0$, there exist $b > 0$, $q_0 > 0$ such that the following holds: if one iteration of the protocol is played with parameters $N, \eta, q \leq q_0$, then*

$$H_{min}^\varepsilon(\Gamma_{EGO}^s | EG) \geq N \cdot \left(\pi\left(\frac{\eta}{\mathbf{v}_g}\right) - \delta \right)$$

and $\varepsilon = \sqrt{2} \cdot 2^{-bqN}$.

The proof follows by finding $q_0, k_0 > 0$ small enough, and M large enough so that for all $q < q_0, k < k_0$:

$$T(\mathbf{v}_G, 2\mathbf{f}_G = 0, \eta, q, k) \geq \pi\left(\frac{\eta}{\mathbf{v}_g}\right) - \frac{\delta}{2}$$

$$E(\mathbf{v}_G, 2\mathbf{f}_G = 0, \eta, q, k) \leq M.$$

Setting $b = \frac{k_0 \cdot \delta}{2M}$ then yields the correct result. We will now find such q_0, k_0 .

We'll start with E . By definition:

$$E = \frac{2}{r_0} = \frac{2}{\min \left\{ \frac{-\mathbf{v}_G}{\pi'(\frac{\eta}{\mathbf{v}_G})}, \frac{1}{qk} \right\}} = 2 \max \left\{ qk, \frac{-\pi'(\frac{\eta}{\mathbf{v}_G})}{\mathbf{v}_G} \right\}$$

So for small enough q_0 and k_0 , specifically:

$$q_0 \cdot k_0 \leq \frac{-\pi'(\frac{\eta}{\mathbf{v}_G})}{\mathbf{v}_G},$$

we have that

$$E = \frac{-2\pi'(\frac{\eta}{\mathbf{v}_G})}{\mathbf{v}_G}$$

We don't know \mathbf{v}_G , but q_0 and k_0 will obey this inequality if:

$$q_0 \cdot k_0 \leq -\pi'(\frac{\eta}{c_v}) \leq \frac{-\pi'(\frac{\eta}{c_v})}{\mathbf{v}_G} \leq \frac{-\pi'(\frac{\eta}{\mathbf{v}_G})}{\mathbf{v}_G}.$$

In this case, we have:

$$E = \frac{-2\pi'(\frac{\eta}{\mathbf{v}_G})}{\mathbf{v}_G} \leq \frac{-2\pi'(\eta)}{c_v} \triangleq M$$

And we have found our M .

Satisfying the condition for T requires a bit more calculations. First we replace \mathbf{v}_G by either 1 or c_v as appropriate, as in the above inequalities. We wish to make the left hand side of the inequality $T(\mathbf{v}_G, 2\mathbf{f}_G = 0, \eta, q, k) \geq \pi(\frac{\eta}{\mathbf{v}_G}) - \frac{\delta}{2}$ smaller, so if we manage to solve that inequality, we will certainly solve the original one.

We therefore look for a q_0, k_0 pair such that for all $q < q_0, k < k_0$,

$$\frac{-1}{r_1 q k} \cdot \max_{t \in [0,1]} \left[\log \left((1-q) 2^{-r_1 q k \cdot \Pi(r_1 q k, t)} + q \left\{ 1 - \left(1 - 2^{-k} \right) c_v^{1+r_1 q k t} \right\} \right) \right] - \frac{\eta}{r_1} \geq \pi\left(\frac{\eta}{\mathbf{v}_G}\right) - \frac{\delta}{2}$$

where $r_1 = \frac{-c_v}{\pi'(\eta)} \leq r_0 = \frac{-\mathbf{v}_G}{\pi'(\frac{\eta}{\mathbf{v}_G})}$.

Note that $\pi(x)$ is a decreasing function in the interval $[0, \frac{1}{2}]$, so $\pi(\frac{\eta}{\mathbf{v}_G}) \leq \pi(\eta)$ as $\mathbf{v}_G \leq 1$. Our inequality will be satisfied if we can satisfy:

$$\frac{-1}{r_1 q k} \cdot \max_{t \in [0,1]} \left[\log \left((1-q) 2^{-r_1 q k \cdot \Pi(r_1 q k, t)} + q \left\{ 1 - \left(1 - 2^{-k} \right) c_v^{1+r_1 q k t} \right\} \right) \right] \geq \pi(\eta) - \frac{\delta}{2} + \frac{\eta}{r_1}$$

This is not easy to do analytically, but numerical calculations can be performed. They show that for each value of δ and η , there is only a small region around $(0,0)$ in which q and k can take values. An example for $\delta = 0.025$, $\eta = 4.2 \cdot 10^{-5}$ can be seen in Figure 2.

It is possible to take any combination of k_0 and q_0 within the specified range. We numerically optimize over such values to find the pair that yields the smallest seed size.

For a given k_0 and using the M previously found, we have:

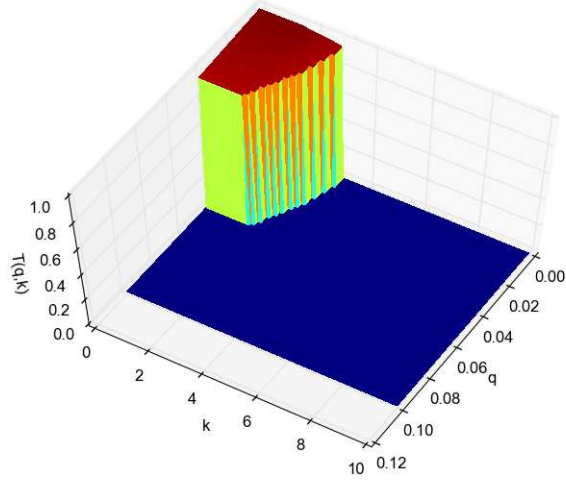


Figure 2: A 3d plot of the value of a clipped T as a function of q and k : the value is zero if it doesn't satisfy the inequality $T(\mathbf{v}_G, 2\mathbf{f}_G, \eta, q, k) \geq \pi(\frac{\eta}{\mathbf{v}_g}) - \frac{\delta}{2}$. As can be seen, only small k and q values are applicable. Here, $\delta = 0.025$ and $\eta = 4.2 \cdot 10^{-5}$

$$b = \frac{k_0 \cdot \delta}{2M} = \frac{k_0 \cdot \delta}{2 \frac{-2\pi'(\eta)}{c_v}} = \frac{k_0 \cdot \delta c_v}{-4\pi'(\eta)}$$

Next, we will find the constants implied in **Corollary I.3**, which shows that there is high min-entropy even conditioned on the input to the device.

Corollary (I.3). *For every $\eta > 0$ and $\delta > 0$, there exist $K, > 0$, $b > 0$, $q_0 > 0$ such that the following holds: if one iteration of the protocol is played with parameters $N, \eta, q \leq q_0$, then*

$$H_{min}^\varepsilon(\Gamma_{EGIO}^s | EGI) \geq N \cdot \left(\pi\left(\frac{\eta}{\mathbf{v}_g}\right) - \delta \right)$$

As in the proof for the corollary, we choose $\delta' = \delta/2$. The associated q'_0 and b' give:

$$\varepsilon' = \sqrt{2} \cdot 2^{-b'qN} = \sqrt{2} \cdot 2^{\frac{k_0 \cdot \delta c_v}{4\pi'(\frac{\eta}{c_v})} Nq}$$

The extra error is, assuming that $q_0 \leq \frac{\delta}{2n}$:

$$e^{-N[\delta/2n - q_0]^2/2}$$

so the total error is now:

$$\begin{aligned} \varepsilon &= \sqrt{2} \cdot 2^{-b'qN} + e^{-N[\delta/2n - q_0]^2/2} \\ &= \sqrt{2} \cdot 2^{-b'qN} + 2^{-\frac{\log e}{2} [\delta/2n - q_0]^2 N} \end{aligned}$$

Assuming b' as before:

$$= \sqrt{2} \cdot 2^{\frac{k_0 \cdot \delta c_v}{4\pi'(\eta)} q N} + 2^{-\frac{\log e}{2} [\delta/2n - q_0]^2 N}$$

We want to bound this from above by a single exponent of the form $K \cdot 2^{-bqN}$. Note that

$$2^{-\frac{\log e}{2} [\delta/2n - q_0]^2 N} \leq 2^{-\frac{\log e}{2} [\delta/2n - q_0]^2 N \frac{q}{q_0}}$$

as $q \leq q_0$ and the exponent would be smaller in magnitude. Taking

$$b = \min \left\{ \frac{k_0 \cdot \delta c_v}{-4\pi'(\eta)}, \frac{\log e}{2} [\delta/2n - q_0]^2 \frac{1}{q_0} \right\},$$

we then have:

$$\begin{aligned} \varepsilon &= \sqrt{2} \cdot 2^{\frac{k_0 \cdot \delta c_v}{-4\pi'(\eta)} q N} + 2^{-\frac{\log 2 \cdot e}{2} [\delta/2n - q_0]^2 N} \\ &\leq \sqrt{2} \cdot 2^{-bqN} + 2^{-bqN} \\ &= (\sqrt{2} + 1) 2^{-bqN} \end{aligned}$$

So $K = \sqrt{2} + 1$.

Next, we find the constants needed for **Corollary I.5**. We restate it here:

Corollary (I.5). *Let $\delta > 0$ be a real number. Then, there exists positive reals K, b, q_0 and η such that the following holds. If Protocol R is executed with parameters N, η, q, G, D , where $q \leq q_0$, then*

$$H_{min}^\varepsilon(\Gamma_{EGIO}^s | EGI) \geq N \cdot (1 - \delta)$$

where $\varepsilon = K \cdot 2^{-bqN}$.

In order to do so, we find an η such that for a given δ , we have:

$$\left| 1 - \pi\left(\frac{\eta}{\mathbf{v}_G}\right) \right| \leq \frac{\delta}{2}$$

And then when we choose $\delta' = \delta/2$ of the original, we are guaranteed to be within the range $[1 - \delta, 1]$, as needed for the corollary. Remembering that $\pi(x)$ is a decreasing function of x , we can set $\mathbf{v}_G = c_v$, and the inequality will still hold. Lets mark $x = \frac{\eta}{c_v}$ and look at the behavior of $\pi(x)$. We want the following inequality to hold:

$$1 - \pi(x) \leq \delta/2$$

Opening up the π function, this reduces to:

$$-x \log x - (1 - x) \log(1 - x) \leq \delta/4$$

This can easily be found numerically, yielding a number $x_0 \in (0, 0.5)$. We then have:

$$\eta_{max} = x_0 \cdot c_v$$

With this we generate an $\eta \in (0, \eta_{max})$ parameter and the constants q_0 , k_0 , K , and b (with a δ value one fourth of the one we used for calculating η , as we had to halve it twice in our proofs). For these parameters, playing one iteration of the expansion protocol will produce a string with min-entropy:

$$H_{min}^\varepsilon(\Gamma_{EGIO}^s | EGI) \geq N \cdot (1 - \delta)$$

where the smoothness is bounded by:

$$\varepsilon = K \cdot 2^{-bqN}$$

Conversely, for a given ε of required smoothness, we have the following constraint on q and N :

$$q \cdot N = -\frac{\log \frac{\varepsilon}{K}}{b} = \frac{\log \frac{K}{\varepsilon}}{b}$$

This will be used when deciding on N and q values for a desired error level.

3.2.2 Random bits for protocol R

Having established a relation between q and N , we can proceed to calculate the number of random bits needed in order to execute protocol R with N games.

Randomness comes into play in two places in protocol R : when deciding on which games we use random inputs instead of dummy zeros, and choosing the actual inputs when this happens. Since we are playing the GHZ game, the latter requires 2 bits for each time we play a real game.

By definition of the bits g_i , generating them requires no more random bits than their Shannon entropy. Combining this with the previous statement, we need

$$2N(-q \log q - (1 - q) \log(1 - q))$$

initial random bits in order to play N games.

3.3 Seed length for the extractor

3.3.1 Quantum secure extractor

Part of the initial randomness needed for one execution of the protocol is the random seed given to the extractor, which we apply on our N bit output that came from playing N games. Based on the paper “Trevisan’s extractor in the presence of quantum side information” [5], we will construct, from bottom up, a suitable extractor. For the purpose of this analysis, we assume that our extractor will operate on N bits which have at least a constant fraction of ε -smooth min-entropy.

Trevisan’s extractor and its quantum security relies on single bit extractors; the ones in [5] use list-decodable codes. All theorems and lemmas in this section are numbered according to [5].

Lemma (C.2). *For every $n \in \mathbb{N}$ and $\delta > 0$, there is code $C_{n,\delta} : \{0,1\}^n \rightarrow \{0,1\}^{\bar{n}}$ that is $(\delta, \frac{1}{\delta^2})$ -list-decodable. Further, $\bar{n} = \text{poly}(n, \frac{1}{\delta})$.*

Guruswami et al. [7] give a construction with $\bar{n} = O(\frac{n}{\delta^4})$; after extracting the constants we have

$$\bar{n} \leq 32 \frac{n}{\delta^4}.$$

Knowing how to construct list-decodable codes, we can use them as extractors:

Theorem (C.3). *Let $C : \{0, 1\}^n \rightarrow \{0, 1\}^{\bar{n}}$ be an (ε, L) list-decodable code. Then*

$$\begin{aligned} C' : \{0, 1\}^n \times [\bar{n}] &\rightarrow \{0, 1\} \\ (x, y) &\mapsto C(x)_y \end{aligned}$$

is a $(\log L + \log \frac{1}{2\varepsilon}, 2\varepsilon)$ -strong extractor.

Notice that as an extractor, the seed given to C' has \bar{n} different inputs, and therefore requires only $t = \log \bar{n}$ bits of randomness.

Combining the two, for any $\varepsilon > 0$ we can build an extractor by putting in $\delta = \frac{\varepsilon}{2}$ into **Lemma C.2**, and using that list-decodable code in **Theorem C.3**. This will give a $(3 \log \frac{1}{\varepsilon}, \varepsilon)$ extractor (actually, we have a $(\log \frac{1}{4\varepsilon^2} + \log \frac{1}{4\varepsilon}, \varepsilon)$ -extractor, but $\log \frac{1}{4\varepsilon^2} + \log \frac{1}{4\varepsilon} = 2 \log \frac{1}{2\varepsilon} + \log \frac{1}{4\varepsilon} = 3 \log \frac{1}{\varepsilon} - 6$, so we certainly have a $(3 \log \frac{1}{\varepsilon}, \varepsilon)$ -extractor as well).

Our extractor requires $t = \log (32 \frac{16n}{\varepsilon^4}) = \log (512 \frac{n}{\varepsilon^4})$ bits of randomness.

Next we will compose 1-bit extractors to create general ones:

Theorem (4.6). *Let $C : \{0, 1\}^n \times \{0, 1\}^t \rightarrow \{0, 1\}$ be a (k, ε) -strong 1-bit extractor with uniform seed, and $S_1, \dots, S_m \subset [d]$ a weak (t, r) -design. Then a Trevisan style extractor composition, $Ext_C : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ gives a $(k + rm + \log \frac{1}{\varepsilon}, 3m\sqrt{\varepsilon})$ -quantum-proof-strong (n, d, m) extractor.*

We are going to use this theorem in order to get a constant rate extractor. Assume that you want an extractor whose output is ε -close to uniform. Then the 1-bit extractor needs to be $\varepsilon' = \frac{\varepsilon^2}{9m^2}$ close to uniform. Set $k = m$ and $r = 2 \frac{k}{m} = 2$. Then the resultant extractor produces output that is ε away from uniform and works with any entropy larger than

$$\begin{aligned} k + rm + \log \frac{1}{\varepsilon'} &= k + 2m + \log \frac{9m^2}{\varepsilon^2} \\ &= k + 2k + 2 \log \frac{3m}{\varepsilon} \\ &= 3k + 2 \log \frac{3k}{\varepsilon} \\ &= 3k + 2 \log \frac{1}{\varepsilon} + 2 \log 3k \\ &\leq 4k + 2 \log \frac{1}{\varepsilon} \end{aligned}$$

Ignoring the $\log \frac{1}{\varepsilon}$ in the min-entropy (it will be small comparable to k), we get that a source with $4k$ min-entropy can give us $m = k$ random bits. Since our N bits have $(1 - \delta)N$ min-entropy, we have $k = \frac{1}{4}(1 - \delta)N$, and also $m = \frac{1}{4}(1 - \delta)N$. Putting this into t , we get:

$$\begin{aligned}
t &= \log(512 \frac{N}{\varepsilon^4}) \\
&= \log(512 \frac{N}{\frac{\varepsilon^8}{9^4 m^8}}) \\
&= \log(512 \cdot 9^4 \frac{Nm^8}{\varepsilon^8}) \\
&= \log(512 \cdot 9^4 \frac{N(\frac{1}{4}(1-\delta)N)^8}{\varepsilon^8}) \\
&= \log(512 \left(\frac{3}{4}(1-\delta)\right)^8 \cdot \frac{N^9}{\varepsilon^8}) \\
&= \log 512 + 8 \log \frac{3}{4}(1-\delta) + \log N + 8 \log \frac{N}{\varepsilon}
\end{aligned}$$

What is d , and how do we get a design? According to **Lemma 5.5** in [5], we can build the desired design with

$$d = t \left\lceil \frac{t}{\ln r} \right\rceil$$

We chose $r = 2$ so this can be effectively written as

$$d \leq \frac{t(t+1)}{\ln 2}$$

3.4 Total randomness for one iteration

Choose a desired distance from uniformity $\varepsilon > 0$, and the two parameters $\delta \in (0, 1)$ and $\eta \leq \eta_{max}$. Find the constants q_0, k_0, b . From the relation $q \cdot N = -\frac{\log \frac{\varepsilon}{K}}{b} = \frac{\log \frac{\sqrt{2}+1}{b\varepsilon}}{b}$ and the fact that $q \leq q_0$, we have a lower bound on N ; pick any N greater than that, and calculate the corresponding q . The total bits of randomness required is then given by the combined result of section 3.2.2 and section 3.3.1:

$$\text{SeedLen} \leq 2N(-q \log q - (1-q) \log(1-q)) + \frac{t(t+1)}{\ln 2}$$

This will generate $\frac{1}{4}(1-\delta)N$ bits which are 2ε close to uniform - one ε is due to the min-entropy smoothness, the other is due to the expander.

This process is only fruitful if the number of generated bits is larger than the number of input bits. This may not be the case for any choice of ε, q , and N ; however, we can always attain this property by increasing N : For a fixed ε , the extractor term grows as $\log^2 N$, while the game generation term grows as $-Nq \log q \propto -\log \frac{1}{N} = \log N$.

Choosing δ and η is not trivial. A small δ value means that the resultant string has higher min-entropy and thus more bits are extracted; however, it also means tighter constraints for q_0 and k_0 . The parameter η affects b, q_0 and k_0 in a non-linear way. Hence, we optimized these parameters numerically for each choice of fixed ε .

4 Multiple iterations and results

The errors for multiple iterations are additive: Using randomness that is ε away from uniformity for an algorithm that expects uniform randomness will add an ε to the output error. In order to keep the error constrained, we must therefore decrease it exponentially (or more) after each time the protocol is played.

The simplest we can do is to cut ε 's value in half after each iteration. This will give no more than 4ε error, and is certainly achievable - the number of output bits grows exponentially, while the increase in the number of bits caused by halving ε is polynomial. This strategy actually overshoots, as the exponential expansion means that there will be many bits left over after each iteration which are not used for the next iteration. Here is an estimation scheme based on the above:

Scheme: The largest seed requirement is imposed by the first iteration. We can minimize the number of excess bits produced in the first iteration, as follows: find a combination of q and N such that the generated number of bits is *just* the amount required for the next iteration (which has ε half the original, so requires more bits). Of course, this number too requires calculation; a simple estimation is achieved assuming that the number of bits generated in the second iteration is in the same proportion to its seed as the number of bits generated in the first iteration is to the initial seed. This scheme means there is no loss of seed in the first iteration, while there may be loss in the next ones; however, it is easy to implement.

Using the first method with an initial error of $0.25 \cdot 10^{-6}$ (to yield a total distance of no more than 10^{-6} away from uniformity) gives an initial seed of less than 715,000 bits. In general, plotting the required number for several different ε values shows a linear relation between the seed length and $\log \frac{1}{\varepsilon}$. This is shown in blue in Figure 3, with a slope of 31328.

A presumably better technique would be to use *all* the generated bits as seed in each consecutive iteration. We set an ε for the first iteration and a target N . Then, for each iteration, we optimize the expression

$$\text{SeedLen} \leq 2N(-q \log q - (1 - q) \log(1 - q)) + \frac{t(t + 1)}{\ln 2}$$

so as to get the smallest ε possible.

The analysis of putting a bound on the resultant $\sum \varepsilon_i$ has not been performed. Further, in order to achieve global optimization one still has to choose an initial N for the first iteration. However, a lower bound for this value can be obtained (for this particular approximation method) by looking at just one iteration: for a given ε , how many initial bits must we use *just* to get back what we invested? As any further iterations just increase the error, and as getting a longer output inherently requires more initial random bits, this gives a bound from below. So running the R protocol once with ε four times the desired value gives us a lower bound. Multiplying ε 's value by four is equivalent to decreasing $\log \frac{1}{\varepsilon}$ by 2. So with the current slope obtained, this can give an improvement of no more than 64,000 bits.

It is interesting to ask which one of the two imposes stronger requirements: generating XOR games, or seeding the extractor. For one iteration, the extractor seed requires $O(\log^2(\frac{N}{\varepsilon}))$ random bits, while the XOR game generation requires $O(Nq \log q)$. However, N , q and ε are connected, and cannot be changed independently. Figure 3 shows the number of bits required as a function of $\log \frac{1}{\varepsilon}$ (applying the same method), assuming that the extractor operates *for free* - it costs us no bits at all to extract. Of course, there are no deterministic extractors, but this gives a bound on the XOR games. It appears that the XOR game generation requires about $2/3$ of the randomness - the ratio between the two slopes is $\frac{\text{with extractor}}{\text{without}} = \frac{31875}{21380} = 1.49$.

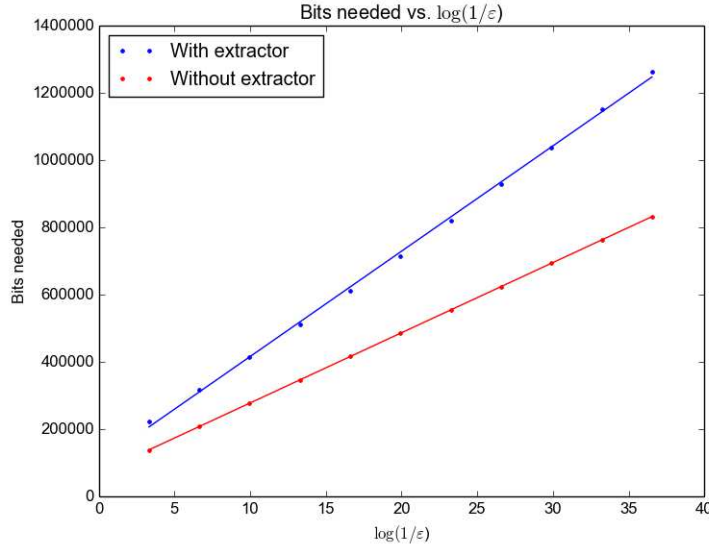


Figure 3: Amount of randomness needed vs. $\log \frac{1}{\varepsilon}$. In blue is the regular scheme; in red is the same scheme but assuming that our extractor is deterministic and requires no random bits.

5 Future work

While we have given a rough upper bound for the amount of randomness needed to “jumpstart” infinite expansion, questions and research directions still remain.

1. What is the optimal XOR game for such a protocol?
2. Is the Coudron-Yuen analysis more or less efficient than the one by Miller and Shi? Can this type of protocol be improved upon? What is the theoretical lower bound for any infinite protocol?
3. How does the number of bits improve if we do not ask for quantum security (security against entanglement with the inner working of the devices), but just want to certify randomness? Are there $O(\log \frac{N}{\varepsilon})$ constant rate strong extractors which can be shown to be quantum secure?

Acknowledgements Renan thanks Scott Aaronson for overview and supervision. We thank Matthew Coudron for clearing things up.

References

- [1] Kai-Min Chung, Yaoyun Shi, and Xiaodi Wu. Physical randomness extractors: Generating random numbers with minimal assumptions. *arXiv:1402.4797*, 2014.
- [2] Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation*. PhD thesis, University of Cambridge, 2006.

- [3] Matthew Coudron, Thomas Vidick, and Henry Yuen. Robust randomness amplifiers: Upper and lower bounds. *arXiv:1305.6626*, 2013.
- [4] Matthew Coudron and Henry Yuen. Infinite randomness expansion and amplification with a constant number of devices. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 427–436, 2014.
- [5] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan’s extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41(4):915–940, 2012.
- [6] Z. Dvir and A. Wigderson. Kakeya Sets, New Mergers, and Old Extractors. *SIAM J. on Computing*, 40(3):778–792, 2011. (Extended abstract appeared in FOCS 2008).
- [7] Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Transactions on Information Theory*, 48:2002, 2000.
- [8] Carl A. Miller and Yaoyun Shi. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. *arXiv:1402.0489*, 2014.
- [9] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by Bell’s theorem. *Nature*, 2010.
- [10] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of chsh games. *arXiv:1209.0448*, 2012.
- [11] Umesh Vazirani and Thomas Vidick. Certifiable quantum dice. *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 61–76, 2012.